

Essential edges in Poisson random hypergraphs

Christina Goldschmidt and James Norris *
Statistical Laboratory, University of Cambridge

February 1, 2008

Abstract

Consider a random hypergraph on a set of N vertices in which, for $1 \leq k \leq N$, a $\text{Poisson}(N\beta_k)$ number of hyperedges is scattered randomly over all subsets of size k . We collapse the hypergraph by running the following algorithm to exhaustion: pick a vertex having a 1-edge and remove it; collapse the hyperedges over that vertex onto their remaining vertices; repeat until there are no 1-edges left. We call the vertices removed in this process *identifiable*. Also any hyperedge all of whose vertices are removed is called *identifiable*. We say that a hyperedge is *essential* if its removal prior to collapse would have reduced the number of identifiable vertices. The limiting proportions, as $N \rightarrow \infty$, of identifiable vertices and hyperedges were obtained in [3]. In this paper, we establish the limiting proportion of essential hyperedges. We also discuss, in the case of a random graph, the relation of essential edges to the 2-core of the graph, the maximal sub-graph with minimal vertex degree 2.

Keywords: Poisson random hypergraphs, essential edges, 2-core, giant component.

1 Introduction

The Poisson random hypergraph model (introduced in [3]) which is the subject of this paper may be considered as a step towards developing random combinatorial structures which one can fit to real-world phenomena. With this aim in mind, the class of hypergraphs in which every edge has the same number of vertices, whilst being a clean and elegant object, may be too narrow to be useful. We find it remarkable that, despite the flexibility afforded by its large number of parameters, the class of Poisson random hypergraphs admits tractable computations for the asymptotic size of key structures.

It is conventional in random combinatorics to impose a uniform rather than a Poisson structure. For large N , this makes little difference so long as one is concerned with “local” random variables, for

*Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK.
Current address for the first author: Laboratoire de Probabilités et Modèles Aléatoires, Université Pierre et Marie Curie (Paris 6), 4, place Jussieu, Boîte courrier 188, 75252 Paris Cedex 05, France; E-mail christina@proba.jussieu.fr

This is a preprint of an article accepted for publication in *Random Structures and Algorithms* Copyright © (2004) Wiley Periodicals, Inc.

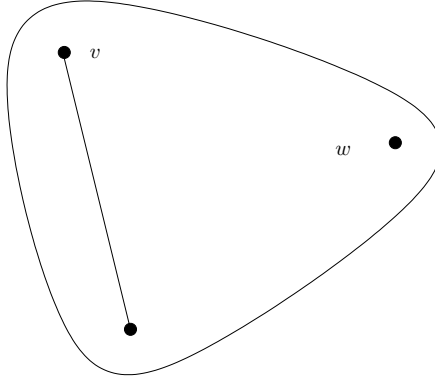


Figure 1: Asymmetry of identifiability

example, the number of edges at a given vertex. A global Poisson structure is natural probabilistically in that it maximizes independence. In contrast, in the uniform case, by specifying an exact total number of hyperedges of a given size, one imposes dependencies at a global level which may be considered unnatural. For many questions, including those addressed in this paper, we would expect to find similar asymptotic behaviour for Poisson and uniform models. This has already been verified by one of us [4] in respect of the numbers of identifiable vertices.

Suppose we have a set of vertices V of size N . Then for $(\beta_k : k \geq 1)$ a sequence of non-negative real numbers, we define a Poisson random hypergraph with parameters $(\beta_k : k \geq 1)$ to be a random map $\Lambda : \mathcal{P}(V) \rightarrow \mathbb{Z}^+$ such that

$$\Lambda(A) \sim \text{Poisson} \left(N\beta_k / \binom{N}{k} \right)$$

whenever $|A| = k$, with $(\Lambda(A), A \subseteq V)$ independent. Then $\Lambda(A)$ is the number of hyperedges over A (so that we are allowing multiple edges over a set). We call 1-edges *patches*. We refer to the case when $\beta_k = 0$ for all $k \geq 3$ as the graph case. Define the generating function $\beta(t) = \sum_{k=1}^{\infty} \beta_k t^k$. Throughout this paper, we will assume that $\sum_{k=1}^{\infty} k\beta_k < \infty$ so that β is C^1 on the interval $[0, 1]$. Define $t^* = \inf\{t \geq 0 : \beta'(t) + \log(1-t) < 0\}$ and note that $t^* \in [0, 1)$. We will assume further that there are no zeros of $\beta'(t) + \log(1-t)$ in $[0, t^*)$. The case where this last condition fails is explored further in [3].

We follow [3], [2] in considering the notion of identifiability. Any vertex with a patch on it is identifiable. Pick such a vertex and delete both the vertex and the patch. Collapse all of the other hyperedges over v down onto their remaining vertices (so that a 3-edge over $\{u, v, w\}$ becomes a 2-edge over $\{u, w\}$, for example). Continue until there are no more patches on the hypergraph. Then the order of this collapse does not affect the set of vertices eventually removed (see [3]), called the set of *identifiable* vertices, which we denote by $V^*(\Lambda)$. A hyperedge is said to be identifiable if all of its vertices are identifiable.

In a hypergraph with no patches there are no identifiable vertices: we say then that a vertex w is *identifiable from* v if it is identifiable in the hypergraph obtained by adding one patch at v . Note that, except in the graph case, this relation between w and v is not symmetric. For example, in Figure 1, w is identifiable from v but v is not identifiable from w . The set of vertices identifiable from v is called the *domain* of v .

We now review some material from [3] and [2] which we will use later.

Theorem 1.1. *Let V_N be the number of identifiable vertices and H_N be the number of identifiable hyperedges in the Poisson random hypergraph on N vertices. Then, for all $\epsilon > 0$,*

$$\limsup_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(|N^{-1}V_N - t^*| > \epsilon) < 0$$

and

$$\limsup_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(|N^{-1}H_N - \beta(t^*) + (1 - t^*) \log(1 - t^*)| > \epsilon) < 0.$$

Thus, V_N/N and H_N/N have limits in probability which are attained exponentially fast.

We recall that the $\text{Borel}(\alpha)$ distribution is the distribution of the total population of a Galton–Watson branching process with $\text{Poisson}(\alpha)$ offspring distribution. That is, if $X \sim \text{Borel}(\alpha)$ then

$$\begin{aligned} \mathbb{P}(X = n) &= e^{-\alpha n} (\alpha n)^{n-1} / n!, \quad n \geq 1 \\ \mathbb{P}(X = \infty) &= p, \end{aligned}$$

with p the root in $(0, 1)$ of $\alpha x + \log(1 - x) = 0$.

Theorem 1.2. *Assume that $\beta_1 = 0$. Let D_N denote the size of the domain of a typical vertex. Then*

$$D_N \xrightarrow{d} D$$

as $N \rightarrow \infty$, where D has the $\text{Borel}(2\beta_2)$ distribution.

The graph case of this result is well known. The domain of a vertex looks like a branching process. This branching process has almost surely finite size if $\beta_2 \leq 1/2$ and is infinite with positive probability if $\beta_2 > 1/2$.

2 Essential edges

We say that a hyperedge is *essential* if removing it reduces the number of identifiable vertices.

Let $\mathcal{E}(N, k)$ be the number of essential k -edges in the Poisson random hypergraph on N vertices. Let $\mathcal{E}(N) = \sum_{k=1}^N \mathcal{E}(N, k)$ be the total number of essential edges. The purpose of this paper is to prove the following law of large numbers:

Theorem 2.1. *As $N \rightarrow \infty$, the following limits hold in probability:*

$$\mathcal{E}(N, k)/N \rightarrow k(1 - t^*)(t^*)^{k-1}\beta_k, \quad k \geq 1, \tag{2.1}$$

$$\mathcal{E}(N)/N \rightarrow -(1 - t^*) \log(1 - t^*). \tag{2.2}$$

Thus, the limit of H_N/N splits into two parts: $\beta(t^*)$ corresponding to non-essential identifiable hyperedges and $-(1 - t^*) \log(1 - t^*) = (1 - t^*)\beta'(t^*)$ corresponding to essential hyperedges.

3 Essential edges in random graphs

In order to provide intuition about essential edges, we consider first the case of a random graph with patches. Asymptotically, the 2-edge-structure of the Poisson random graph behaves in the same way as that of the more-commonly studied binomial model $\mathbb{G}(N, p)$, with $p = 2\beta_2/N$ (see, for example, Bollobás [1] or Janson, Łuczak and Ruciński [6]). We give a simple calculation for essential patches and then a heuristic derivation for essential 2-edges based on known results for the 2-core.

In the random graph with patches, a vertex is identifiable if and only if its component has a patch on some vertex. So a patch is essential if and only if it is the only patch on a component. This enables us to prove part of our limiting result concerning the expected number of essential patches in an elementary way. Fix a vertex v and let D_N be the size of the domain of v . By Theorem 1.2, $D_N \xrightarrow{d} D$. We note that the limiting Borel($2\beta_2$) distribution for D has (possibly degenerate) probability generating function $F(s)$ which is the solution to

$$F(s) = s \exp(2\beta_2(F(s) - 1))$$

in the range $0 \leq s \leq 1$ (see Harris [5] p.32). Let A denote the event that v has a patch and this patch is essential. Then

$$\mathbb{P}(A) = \mathbb{E}[\mathbb{P}(A|D_N)] = \mathbb{E}[\beta_1 e^{-\beta_1 D_N}] \rightarrow \beta_1 F(e^{-\beta_1}).$$

Now,

$$F(e^{-\beta_1}) = \exp(-\beta_1 + 2\beta_2(F(e^{-\beta_1}) - 1)).$$

But $F(e^{-\beta_1}) = 1 - t^*$ is a solution to this equation and so

$$\mathbb{P}(A) \rightarrow \beta_1(1 - t^*).$$

Hence, $\mathbb{E}[\mathcal{E}(N, 1)/N]$ converges to $\beta_1(1 - t^*)$.

A 2-edge in a random graph with patches is essential if and only if removing it splits its component into two disconnected components, exactly one of which has a patch on it. In particular, edges in cycles are not essential; an edge in a path between two patches is not essential; edges in components with no patches cannot be essential (they are not even identifiable). All edges in a tree-component with a single patch are essential.

We now describe a connection with the 2-core of a random graph, that is the maximal subgraph with minimum degree 2. (The 2-core consists of all vertices on cycles and on paths between cycles.) The 2-core of a random graph is only of size $\Theta(N)$ if the random graph is super-critical (i.e. $2\beta_2 > 1$), so we work with that case. This section is intended to provide orientation for our general results in a context which may be more familiar to some readers. We do not attempt to provide a fully rigorous discussion.

It is known (see Pittel [7]) that the giant component consists of its own 2-core and a *mantle* of trees, each sprouting from a different vertex of the 2-core. The proportion of vertices in the giant 2-core is a.s. $\theta - 2\beta_2\theta(1 - \theta)$ and the proportion of vertices in the mantle is $2\beta_2\theta(1 - \theta)$, where $\theta = \inf\{t \geq 0 : 2\beta_2 t + \log(1 - t) < 0\}$. Note that $\theta = t^*$ when we have $\beta_1 = 0$ i.e. no patches. We can imagine instead that we have an $o(N)$ number of patches, enough to ensure that a.s. one lands on the giant component. Then the process of identification basically picks out the giant component (and a few other smaller tree-like components which we may neglect because they are of size at most

$\mathcal{O}(\log N)$). Thus, the proportion of identifiable vertices is t^* which corresponds to the size of the giant component. As edges in cycles cannot be essential, the essential edges in the giant component must almost all be found in the mantle. Because the mantle is like a forest, the number of edges and vertices in it are approximately equal and so the number of essential edges scaled by $1/N$ is the same as the size of the mantle scaled by $1/N$, that is $2\beta_2 t^*(1 - t^*)$.

4 Convergence of expectations

Let $\Lambda \mathbb{1}_{\{A\}^c}$ be the Poisson random hypergraph Λ with any hyperedges over the set A removed. We say that a set $A \subseteq V$ is *essential* for Λ if $\Lambda(A) = 1$ and $V^*(\Lambda \mathbb{1}_{\{A\}^c}) \neq V^*(\Lambda)$. Thus a set is essential if and only if it has an essential hyperedge over it.

Proposition 4.1. *A set A is essential if and only if $\Lambda(A) = 1$ and $|A \setminus V^*(\Lambda \mathbb{1}_{\{A\}^c})| = 1$.*

Proof. It is clear that a set cannot be essential if it has no hyperedges over it. Also, it cannot be essential if it has more than one hyperedge over it. Recall that the order of deletion does not affect the set of identifiable edges. Suppose $|A \setminus V^*(\Lambda \mathbb{1}_{\{A\}^c})| = 0$. Then everything in A is identifiable without the hyperedge over A and so if we were to re-introduce the hyperedge, it would not be essential. If $|A \setminus V^*(\Lambda \mathbb{1}_{\{A\}^c})| \geq 2$ then replacing the hyperedge over A will not make the two or more vertices identifiable and so A cannot have been essential. There remains the case $\Lambda(A) = 1$ with $|A \setminus V^*(\Lambda \mathbb{1}_{\{A\}^c})| = 1$, when A is obviously essential. \square

Lemma 4.2. *For any fixed $k \geq 1$, we have $\mathbb{E}[\mathcal{E}(N, k)/N] \rightarrow k(1 - t^*)(t^*)^{k-1}\beta_k$ as $N \rightarrow \infty$.*

Proof. The numbers of hyperedges on distinct subsets of V are independent and so $\Lambda \mathbb{1}_{\{A\}^c}$ has the same law as Λ conditioned on $\Lambda(A) = 0$. Fix $k \geq 1$ and choose A with $|A| = k$. Let $q^{(1)}(N, k)$ be the probability that A is essential. Then, by Proposition 4.1,

$$\begin{aligned} q^{(1)}(N, k) &= \mathbb{P}(\Lambda(A) = 1, |A \setminus V^*(\Lambda \mathbb{1}_{\{A\}^c})| = 1) \\ &= \mathbb{P}(\Lambda(A) = 1) \mathbb{P}(|A \setminus V^*(\Lambda \mathbb{1}_{\{A\}^c})| = 1) \\ &= \mathbb{P}(\Lambda(A) = 1) \mathbb{P}(|A \setminus V^*| = 1 | \Lambda(A) = 0) \\ &= \mathbb{P}(\Lambda(A) = 1) \frac{\mathbb{P}(|A \setminus V^*| = 1)}{\mathbb{P}(\Lambda(A) = 0)} \quad \text{as } |A \setminus V^*| = 1 \text{ implies } \Lambda(A) = 0 \\ &= \frac{N\beta_k}{\binom{N}{k}} \mathbb{P}(|A \setminus V^*| = 1). \end{aligned}$$

Hence,

$$\mathbb{E}[\mathcal{E}(N, k)/N] = \binom{N}{k} q^{(1)}(N, k)/N = \beta_k \mathbb{P}(|A \setminus V^*| = 1). \quad (4.1)$$

Now, if $|A \setminus V^*| = 1$, then A must contain $k - 1$ identifiable vertices and one non-identifiable vertex

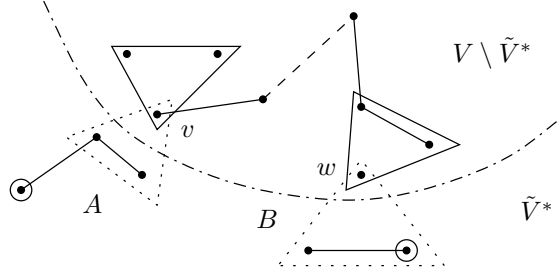


Figure 2: Example of two essential edges. The dot-dashed curve indicates the boundary between \tilde{V}^* and $V \setminus \tilde{V}^*$.

and symmetry implies that, given $|V^*|$, all vertices are equally likely to be identifiable. So,

$$\begin{aligned} \mathbb{P}(|A \setminus V^*| = 1 \mid |V^*|) &= \frac{\binom{|V^*|}{k-1} (N - |V^*|)}{\binom{N}{k}} \\ &= k \left(1 - \frac{|V^*|}{N}\right) \frac{|V^*|}{(N-1)} \frac{(|V^*| - 1)}{(N-2)} \dots \frac{(|V^*| - k + 2)}{(N - k + 1)} \\ &\xrightarrow{p} k(1 - t^*)(t^*)^{k-1} \end{aligned}$$

since $|V^*|/N \xrightarrow{p} t^*$. Because $0 \leq \mathbb{P}(|A \setminus V^*| = 1 \mid |V^*|) \leq 1$, it follows by bounded convergence that

$$\mathbb{P}(|A \setminus V^*| = 1) = \mathbb{E}[\mathbb{P}(|A \setminus V^*| = 1 \mid |V^*|)] \rightarrow k(1 - t^*)(t^*)^{k-1}$$

as $N \rightarrow \infty$. Thus, by (4.1), $\mathbb{E}[\mathcal{E}(N, k)/N] \rightarrow k(1 - t^*)(t^*)^{k-1}\beta_k$. \square

5 Asymptotic independence

The key point is to show that the events $\{A \text{ is essential}\}$ and $\{B \text{ is essential}\}$ are asymptotically independent, for distinct sets A and B . Once we have done this, Theorem 2.1 can be proved in much the same way as the weak law of large numbers for independent random variables.

We now proceed as in the proof of Lemma 4.2. Suppose that $|A| = |B| = k$ where $A \neq B$. Let $q^{(2)}(N, k)$ be the probability that both A and B are essential. Then we have

$$q^{(2)}(N, k) = \mathbb{P}(\Lambda(A) = 1, \Lambda(B) = 1, |A \setminus V^*(\Lambda \mathbb{1}_{\{A\}^c})| = 1, |B \setminus V^*(\Lambda \mathbb{1}_{\{B\}^c})| = 1).$$

Now, if we could deal with $\Lambda \mathbb{1}_{\{A, B\}^c}$ instead of $\Lambda \mathbb{1}_{\{A\}^c}$ and $\Lambda \mathbb{1}_{\{B\}^c}$ then we would be able to proceed easily by saying that $\Lambda \mathbb{1}_{\{A, B\}^c}$ has the same law as Λ conditioned on $\Lambda(A) = 0$ and $\Lambda(B) = 0$. However, we must then deal explicitly with the cases where A contributes towards the identifiability of B , or vice versa. For ease of notation, write \tilde{V}^* for $V^*(\Lambda \mathbb{1}_{\{A, B\}^c})$. We commence with some examples. In Figure 2, $|A \setminus \tilde{V}^*| = 1$ and $|B \setminus \tilde{V}^*| = 1$. Both A and B are essential as long as v is not in the domain of w in $V \setminus \tilde{V}^*$ and w is not in the domain of v in $V \setminus \tilde{V}^*$ (if the dashed edge is present then w is identifiable from v and so B is not essential). In Figure 3, both A and B

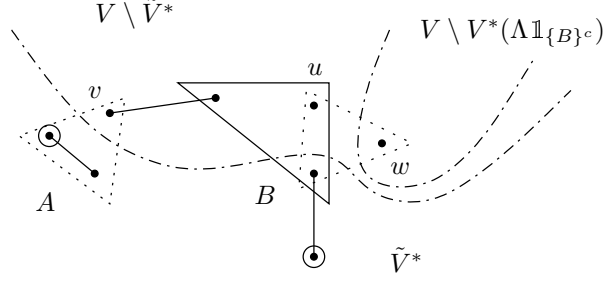


Figure 3: Example of two essential edges. The dot-dashed curves indicate the boundaries between \tilde{V}^* and $V \setminus \tilde{V}^*$ and between $V \setminus \tilde{V}^*$ and $V \setminus V^*(\Lambda \mathbb{1}_{\{B\}^c})$.

are essential but $|B \setminus \tilde{V}^*| = 2$. The important point here is that precisely one element u of $B \setminus \tilde{V}^*$ is identifiable from v and the other is not.

In general, in order to have $|A \setminus V^*(\Lambda \mathbb{1}_{\{A\}^c})| = 1$ and $|B \setminus V^*(\Lambda \mathbb{1}_{\{B\}^c})| = 1$ we must always have either $|A \setminus \tilde{V}^*| = 1$ or $|B \setminus \tilde{V}^*| = 1$ because otherwise none of $V \setminus \tilde{V}^*$ would be identifiable when we reintroduce the edges over A and B . Suppose, without loss of generality, that $A \setminus \tilde{V}^* = \{v\}$. Then we must also have that all but one of the vertices in $B \setminus \tilde{V}^*$ are in the domain of v .

Lemma 5.1. *Let $\tilde{\mathcal{D}}(v)$ be the domain of v in the collapsed hypergraph on $V \setminus \tilde{V}^*$ and let $\mathcal{D}(v)$ be the domain of v in the collapsed hypergraph on $V \setminus V^*$. Let $\{w\} = B \setminus \tilde{V}^*$ when $|B \setminus \tilde{V}^*| = 1$. Assume that $v \neq w$. Then,*

$$\begin{aligned} & \mathbb{P}(\Lambda(A) = 1, \Lambda(B) = 1, |A \setminus V^*(\Lambda \mathbb{1}_{\{A\}^c})| = 1, |B \setminus V^*(\Lambda \mathbb{1}_{\{B\}^c})| = 1) \\ &= \left(\frac{N\beta_k}{\binom{N}{k}} \right)^2 \left[\mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = 1, v \notin \mathcal{D}(w), w \notin \mathcal{D}(v)) \right. \\ & \quad \left. + 2 \sum_{i=2}^k \mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = i, |\mathcal{D}(v) \cap B \setminus V^*| = i-1, \Lambda(A) = 0, \Lambda(B) = 0) \right]. \quad (5.1) \end{aligned}$$

Proof. We have

$$\begin{aligned} & \mathbb{P}(\Lambda(A) = 1, \Lambda(B) = 1, |A \setminus V^*(\Lambda \mathbb{1}_{\{A\}^c})| = 1, |B \setminus V^*(\Lambda \mathbb{1}_{\{B\}^c})| = 1) \\ &= \mathbb{P}(\Lambda(A) = 1, \Lambda(B) = 1, |A \setminus \tilde{V}^*| = 1, |B \setminus \tilde{V}^*| = 1, v \notin \tilde{\mathcal{D}}(w), w \notin \tilde{\mathcal{D}}(v)) \\ & \quad + \sum_{i=2}^k \mathbb{P}(\Lambda(A) = 1, \Lambda(B) = 1, |A \setminus \tilde{V}^*| = 1, |B \setminus \tilde{V}^*| = i, |\tilde{\mathcal{D}}(v) \cap B \setminus \tilde{V}^*| = i-1) \\ & \quad + \sum_{i=2}^k \mathbb{P}(\Lambda(A) = 1, \Lambda(B) = 1, |B \setminus \tilde{V}^*| = 1, |A \setminus \tilde{V}^*| = i, |\tilde{\mathcal{D}}(w) \cap A \setminus \tilde{V}^*| = i-1) \\ &= \mathbb{P}(\Lambda(A) = 1) \mathbb{P}(\Lambda(B) = 1) \left[\mathbb{P}(|A \setminus \tilde{V}^*| = 1, |B \setminus \tilde{V}^*| = 1, v \notin \tilde{\mathcal{D}}(w), w \notin \tilde{\mathcal{D}}(v)) \right. \\ & \quad \left. + 2 \sum_{i=2}^k \mathbb{P}(|A \setminus \tilde{V}^*| = 1, |B \setminus \tilde{V}^*| = i, |\tilde{\mathcal{D}}(v) \cap B \setminus \tilde{V}^*| = i-1) \right] \end{aligned}$$

$$\begin{aligned}
&= \left(\frac{N\beta_k}{\binom{N}{k}} \exp(-N\beta_k / \binom{N}{k}) \right)^2 \\
&\quad \times \left[\mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = 1, v \notin \mathcal{D}(w), w \notin \mathcal{D}(v) | \Lambda(A) = 0, \Lambda(B) = 0) \right. \\
&\quad \left. + 2 \sum_{i=2}^k \mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = i, |\mathcal{D}(v) \cap B \setminus V^*| = i-1 | \Lambda(A) = 0, \Lambda(B) = 0) \right] \\
&= \left(\frac{N\beta_k}{\binom{N}{k}} \exp(-N\beta_k / \binom{N}{k}) \right)^2 \left[\frac{\mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = 1, v \notin \mathcal{D}(w), w \notin \mathcal{D}(v))}{\mathbb{P}(\Lambda(A) = 0) \mathbb{P}(\Lambda(B) = 0)} \right. \\
&\quad \left. + 2 \sum_{i=2}^k \frac{\mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = i, |\mathcal{D}(v) \cap B \setminus V^*| = i-1, \Lambda(A) = 0, \Lambda(B) = 0)}{\mathbb{P}(\Lambda(A) = 0) \mathbb{P}(\Lambda(B) = 0)} \right]
\end{aligned}$$

as $|A \setminus V^*| = 1, |B \setminus V^*| = 1$ implies $\Lambda(A) = 0, \Lambda(B) = 0$. But then cancellation means that the last line is equal to

$$\begin{aligned}
&\left(\frac{N\beta_k}{\binom{N}{k}} \right)^2 \left[\mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = 1, v \notin \mathcal{D}(w), w \notin \mathcal{D}(v)) \right. \\
&\quad \left. + 2 \sum_{i=2}^k \mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = i, |\mathcal{D}(v) \cap B \setminus V^*| = i-1, \Lambda(A) = 0, \Lambda(B) = 0) \right],
\end{aligned}$$

as required. \square

For the moment, we will assume that

$$\mathbb{P}(v \notin \mathcal{D}(w), w \notin \mathcal{D}(v)) \rightarrow 1 \quad (5.2)$$

$$\mathbb{P}(|\mathcal{D}(v) \cap B \setminus V^*| \geq 1) \rightarrow 0 \quad (5.3)$$

as $N \rightarrow \infty$ which, in particular, means that we may discard the second term in (5.1). These results will be proved later and reflect the fact that the collapse algorithm will not die out while the vertex domains remain super-critical. Finally, as we will also show later,

$$\mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = 1) \rightarrow (k(t^*)^{k-1}(1-t^*))^2$$

and so

$$q^{(2)}(N, k) \sim \left(\frac{N\beta_k}{\binom{N}{k}} k(t^*)^{k-1}(1-t^*) \right)^2 \sim q^{(1)}(N, k)^2,$$

where here \sim means that the ratio of the left and right sides tends to 1.

6 Proof of Theorem 2.1

We wish to prove the statements (5.2) and (5.3). As a first step, we prove

Lemma 6.1. *Let $\gamma(t) = (1 - t)\beta''(t)$. Then*

$$\gamma(t^*) \leq 1.$$

Proof. For $t \in [0, 1]$, define $f(t) = 1 - t - \exp(-\beta'(t))$. Then

$$f'(t) = \beta''(t) \exp(-\beta'(t)) - 1.$$

Now $t^* = \inf\{t \geq 0 : f(t) < 0\}$ and so $f'(t^*) = (1 - t^*)\beta''(t^*) - 1 \leq 0$. Hence result. \square

As we shall soon see, the 2-edge parameter for the collapsed hypergraph is approximately $\frac{1}{2}\gamma(t^*)$ and so this lemma says that the collapsed hypergraph is nearly sub-critical.

Lemma 6.2. *Suppose that $v, w \notin V^*$ are chosen uniformly at random. Recall that $\mathcal{D}(v)$ is the domain of v in the collapsed hypergraph on $V \setminus V^*$. Then*

$$\mathbb{P}(w \in \mathcal{D}(v)) \rightarrow 0$$

as $N \rightarrow \infty$.

Proof. Let Λ^{V^*} be the hypergraph obtained from Λ by removing all of the vertices in V^* , so that Λ^{V^*} is the collapsed hypergraph. Clearly, for $v \in V \setminus V^*$,

$$\Lambda^{V^*}(\{v\}) = 0$$

as v is not identifiable. We need to find the distribution of $\Lambda^{V^*}(A)$ for all $A \in V \setminus V^*$. Suppose that each set in $\mathcal{P}(V)$ has a corresponding card which gives the number of hyperedges on it. Initially, we place the cards face down, so that we know the sets they represent but not the numbers of hyperedges. Consider the following slightly different way of looking at the process of identification. First turn over the cards corresponding to all the singleton sets; this tells us which vertices have patches on them. Write a list, \mathcal{L} of the vertices with patches. Now proceed recursively. Pick any set with all but one of its vertices in \mathcal{L} and turn its card over. Add the last vertex to \mathcal{L} if there is an edge over the set; if there is no edge, \mathcal{L} remains unchanged. Repeat. The process terminates when we have run out of sets with all but one of their vertices in \mathcal{L} , so that $\mathcal{L} = V^*$ is the set of identifiable vertices. Discard all of the cards which have been turned over, as well as any corresponding to subsets of V^* . Then there remain only the cards corresponding to sets with at least two vertices in $V \setminus V^*$. As each carried an independent random variable on its face at the start and we have not turned any of them over in the process of identification, the random variables must remain independent of one another and of what we have seen of the hypergraph on V^* . Thus, conditional on V^* with $|V^*| = n$, we have that for $A \subseteq V \setminus V^*$ with $|A| = j \geq 2$,

$$\Lambda^{V^*}(A) = \sum_{\substack{B \supseteq A, \\ B \setminus V^* = A}} \Lambda(B) \sim \text{Poisson} \left(N \sum_{i=0}^n \beta_{i+j} \binom{n}{i} / \binom{N}{i+j} \right)$$

and these random variables are independent. So Λ^{V^*} is a new Poisson random hypergraph on $N - n$ vertices and with parameters $\beta_k(n, N)$ where

$$\beta_k(n, N) = \begin{cases} 0 & \text{if } k = 1 \\ \frac{N}{N-n} \binom{N-n}{k} \sum_{i=0}^n \beta_{i+k} \binom{n}{i} / \binom{N}{i+k} & \text{if } k \geq 2. \end{cases}$$

Choose $\rho \in (t^*, 1)$. By Lemma 6.1 of Darling and Norris [3], we obtain

$$|\beta_2(n, N) - \frac{1}{2}\gamma(\frac{n}{N})| \leq C(\log N)^2/N$$

for some constant $C < \infty$ and all $n \leq [\rho N]$.

Furthermore, for each k , the number of k -edges in the new hypergraph is certainly bounded by the total number of hyperedges in the original hypergraph, which had a $\text{Poisson}(N\beta(1))$ distribution. Thus,

$$\beta_k(n, N) \leq \frac{N\beta(1)}{N-n} \leq \frac{\beta(1)}{1-\rho}$$

for all $k \geq 2$ and all $n \leq [\rho N]$.

Now let C_t have the distribution of the size of the domain of a vertex in a Poisson random hypergraph with parameters $\beta_1 = 0$, $\beta_2 = \frac{1}{2}\gamma(t) + C(\log N)^2/N$ and $\beta_k = \beta(1)/(1-\rho)$ for $k \geq 3$. Then by Theorem 1.2 and an obvious comparison argument,

$$C_t \xrightarrow{d} \text{Borel}(\gamma(t))$$

as $N \rightarrow \infty$. Also, if $|V^*| = Nt$ for $t < \rho$ then $|\mathcal{D}(v)|$ is stochastically dominated by C_t .

For any $\delta > 0$, choose $\epsilon > 0$ small enough and $R < \infty$ large enough that, firstly, $t^* + \epsilon < \rho$ and, secondly, that

$$\mathbb{P}(\text{Borel}(\gamma(t^* + \epsilon)) \leq R) \geq 1 - \delta$$

(we can do this because $\gamma(t^*) \leq 1$ and γ is continuous). Hence,

$$\begin{aligned} \mathbb{P}(w \in \mathcal{D}(v)) &= \mathbb{P}\left(w \in \mathcal{D}(v), \left|\frac{|V^*|}{N} - t^*\right| \leq \epsilon\right) + \mathbb{P}\left(w \in \mathcal{D}(v), \left|\frac{|V^*|}{N} - t^*\right| > \epsilon\right) \\ &\leq \mathbb{P}\left(w \in \mathcal{D}(v), |\mathcal{D}(v)| \leq R, \left|\frac{|V^*|}{N} - t^*\right| \leq \epsilon\right) \\ &\quad + \mathbb{P}\left(|\mathcal{D}(v)| > R, \left|\frac{|V^*|}{N} - t^*\right| \leq \epsilon\right) + \mathbb{P}\left(\left|\frac{|V^*|}{N} - t^*\right| > \epsilon\right) \\ &\leq \frac{R}{N(1-t^*-\epsilon)} + \mathbb{P}(C_{t^*+\epsilon} > R) + \mathbb{P}\left(\left|\frac{|V^*|}{N} - t^*\right| > \epsilon\right) \\ &\rightarrow \mathbb{P}(\text{Borel}(\gamma(t^* + \epsilon)) > R) \quad \text{as } N \rightarrow \infty \end{aligned}$$

But this last quantity is less than δ and so we are done. \square

Finally, we give a technical lemma.

Lemma 6.3. *Suppose that for $k \geq 1$ and $N \geq 1$, $X_{k,N}$ are non-negative random variables satisfying*

$$\frac{1}{N}X_{k,N} \xrightarrow{p} x_k \tag{6.1}$$

as $N \rightarrow \infty$, for all k , where

$$\sum_{k=1}^{\infty} x_k < \infty. \tag{6.2}$$

Suppose in addition that for each k there exists y_k such that

$$\mathbb{E} \left[\frac{1}{N} X_{k,N} \right] \leq y_k \quad (6.3)$$

for all N with

$$\sum_{k=1}^{\infty} y_k < \infty. \quad (6.4)$$

Then

$$\frac{1}{N} \sum_{k=1}^N X_{k,N} \xrightarrow{p} \sum_{k=1}^{\infty} x_k$$

as $N \rightarrow \infty$.

Proof. Let $\epsilon, \delta > 0$. By (6.2) and (6.4), we can find k_0 sufficiently large that

$$\sum_{k=k_0+1}^{\infty} x_k + \sum_{k=k_0+1}^{\infty} y_k < \delta\epsilon/4. \quad (6.5)$$

Moreover, as k_0 is fixed and finite, by (6.1) we have that $N^{-1} \sum_{k=1}^{k_0} X_{k,N} \xrightarrow{p} \sum_{k=1}^{k_0} x_k$ and so, for N sufficiently large,

$$\mathbb{P} \left(\left| \frac{1}{N} \sum_{k=1}^{k_0} X_{k,N} - \sum_{k=1}^{k_0} x_k \right| > \epsilon/2 \right) < \delta/2. \quad (6.6)$$

Now,

$$\begin{aligned} & \mathbb{P} \left(\left| \frac{1}{N} \sum_{k=1}^N X_{k,N} - \sum_{k=1}^{\infty} x_k \right| > \epsilon \right) \\ & \leq \mathbb{P} \left(\left| \frac{1}{N} \sum_{k=1}^{k_0} X_{k,N} - \sum_{k=1}^{k_0} x_k \right| > \epsilon/2 \right) + \mathbb{P} \left(\left| \frac{1}{N} \sum_{k=k_0+1}^N X_{k,N} - \sum_{k=k_0+1}^{\infty} x_k \right| > \epsilon/2 \right) \\ & \leq \mathbb{P} \left(\left| \frac{1}{N} \sum_{k=1}^{k_0} X_{k,N} - \sum_{k=1}^{k_0} x_k \right| > \epsilon/2 \right) + \frac{2}{\epsilon} \left\{ \sum_{k=k_0+1}^{\infty} y_k + \sum_{k=k_0+1}^{\infty} x_k \right\} \end{aligned}$$

by Markov's inequality and (6.3). But by (6.5) and (6.6) this last expression is less than δ for all N sufficiently large. \square

Now we are ready to prove Theorem 2.1.

Proof of Theorem 2.1. Firstly observe that (2.2) follows from (2.1) by applying Lemma 6.3 with $X_{k,N} = \mathcal{E}(N, k)$, $x_k = k(1 - t^*)(t^*)^{k-1} \beta_k$ and $y_k = \beta_k$ (the number of essential k -edges per vertex is bounded in expectation by the total number of k -edges per vertex), where we have $\sum_{k=1}^{\infty} k(1 - t^*)(t^*)^{k-1} \beta_k < \infty$ and $\sum_{k=1}^{\infty} \beta_k < \infty$ by assumption.

It remains to prove (2.1). Take $\epsilon > 0$ and fix $k \geq 1$. Note that

$$\frac{1}{N} \mathcal{E}(N, k) = \frac{1}{N} \sum_{\substack{A \subseteq V_i \\ |A|=k}} \mathbb{1}_{\{A \text{ is essential}\}}.$$

By Chebyshev's inequality, for N sufficiently large,

$$\begin{aligned}
& \mathbb{P} \left(\left| \frac{1}{N} \sum_{\substack{A \subseteq V, \\ |A|=k}} \mathbb{1}_{\{A \text{ is essential}\}} - (1-t^*)k\beta_k(t^*)^{k-1} \right| > \epsilon \right) \\
& \leq \frac{1}{\epsilon^2} \mathbb{E} \left[\left(\frac{1}{N} \sum_{\substack{A \subseteq V, \\ |A|=k}} \mathbb{1}_{\{A \text{ is essential}\}} - (1-t^*)k\beta_k(t^*)^{k-1} \right)^2 \right] \\
& = \frac{1}{\epsilon^2} \left\{ \frac{1}{N^2} \sum_{\substack{A, B \subseteq V, A \neq B \\ |A|=|B|=k}} \mathbb{P}(A \text{ is essential}, B \text{ is essential}) - ((1-t^*)k\beta_k(t^*)^{k-1})^2 \right\} \\
& \quad + \frac{1}{\epsilon^2} \left\{ \frac{1}{N^2} \mathbb{E}[\mathcal{E}(N, k)] + 2(1-t^*)k\beta_k(t^*)^{k-1} \left[(1-t^*)k\beta_k(t^*)^{k-1} - \frac{1}{N} \mathbb{E}[\mathcal{E}(N, k)] \right] \right\},
\end{aligned}$$

where the second term in braces tends to 0 as $N \rightarrow \infty$, by Lemma 4.2. By Lemma 5.1,

$$\begin{aligned}
& \frac{1}{N^2} \sum_{\substack{A, B \subseteq V, A \neq B \\ |A|=|B|=k}} \mathbb{P}(A \text{ is essential}, B \text{ is essential}) \\
& = \frac{1}{N^2} \sum_{\substack{A, B \subseteq V, A \neq B \\ |A|=|B|=k}} \frac{N^2 \beta_k^2}{\binom{N}{k}^2} \left\{ \mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = 1, v \notin \mathcal{D}(w), w \notin \mathcal{D}(v)) \right. \\
& \quad \left. + 2 \sum_{i=2}^k \mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = i, |\mathcal{D}(v) \cap B \setminus V^*| = i-1, \Lambda(A) = 0, \Lambda(B) = 0) \right\}.
\end{aligned}$$

Now, by Lemma 6.2, $\mathbb{P}(v \notin \mathcal{D}(w), w \notin \mathcal{D}(v)) \rightarrow 1$ and $\mathbb{P}(|\mathcal{D}(v) \cap B \setminus V^*| \geq 1) \rightarrow 0$ as $N \rightarrow \infty$. Thus, we are really interested in

$$\begin{aligned}
& \frac{\beta_k^2}{\binom{N}{k}^2} \sum_{\substack{A, B \subseteq V, A \neq B \\ |A|=|B|=k}} \mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = 1) \\
& = \frac{\beta_k^2}{\binom{N}{k}^2} \sum_{\substack{A, B \subseteq V, A \cap B = \emptyset \\ |A|=|B|=k}} \mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = 1) \\
& \quad + \frac{\beta_k^2}{\binom{N}{k}^2} \sum_{i=1}^{k-1} \sum_{\substack{A, B \subseteq V, |A \cap B|=i \\ |A|=|B|=k}} \mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = 1). \tag{6.7}
\end{aligned}$$

In the case where A and B are disjoint, by a similar argument to that used in the proof of Lemma 4.2, we have

$$\mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = 1) = \mathbb{E} \left[\frac{(N - |V^*|)(N - |V^*| - 1) \binom{|V^*|}{k-1} \binom{|V^*| - k + 1}{k-1}}{\binom{N}{k} \binom{N-k}{k}} \right].$$

If $|A \cap B| = i$ for $1 \leq i \leq k-1$, then the two non-identifiable vertices must lie outside the intersection (otherwise they are the same vertex and so $v = w$). Thus,

$$\begin{aligned} & \mathbb{P}(|A \setminus V^*| = 1, |B \setminus V^*| = 1) \\ &= \mathbb{E} \left[\frac{(N - |V^*|) \binom{|V^*|}{k-i-1} \binom{|V^*|-k+i+1}{i} (N - |V^*| - 1) \binom{|V^*|-k+1}{k-i-1}}{\binom{N}{k} \binom{k}{i} \binom{N-k}{k-i}} \right]. \end{aligned}$$

So (6.7) is equal to

$$\begin{aligned} & \frac{\beta_k^2}{\binom{N}{k}^2} \mathbb{E} \left[(N - |V^*|)(N - |V^*| - 1) \binom{|V^*|}{k-1} \binom{|V^*|-k+1}{k-1} \right] \\ &+ \frac{\beta_k^2}{\binom{N}{k}^2} \sum_{i=1}^{k-1} \mathbb{E} \left[(N - |V^*|) \binom{|V^*|}{k-i-1} \binom{|V^*|-k+i+1}{i} (N - |V^*| - 1) \binom{|V^*|-k+1}{k-i-1} \right]. \end{aligned}$$

As $\frac{|V^*|}{N} \rightarrow t^*$ in probability, the first of these two terms converges to

$$(k\beta_k(1-t^*)(t^*)^{k-1})^2,$$

by bounded convergence. It remains to show that the second term converges to 0. We have

$$\begin{aligned} & \sum_{i=1}^{k-1} \mathbb{E} \left[\frac{(N - |V^*|) \binom{|V^*|}{k-i-1} \binom{|V^*|-k+i+1}{i} (N - |V^*| - 1) \binom{|V^*|-k+1}{k-i-1}}{\binom{N}{k}^2} \right] \\ & \leq \sum_{i=1}^{k-1} \frac{(k!)^2}{i!((k-i-1)!)^2(N-k+i+1)!} \\ & \rightarrow 0 \end{aligned}$$

as $N \rightarrow \infty$. The result follows. \square

References

- [1] B. Bollobás, *Random Graphs*, Academic Press, New York, 1985.
- [2] R.W.R. Darling, D.A. Levin, J.R. Norris, Continuous and discontinuous phase transitions in hypergraph processes, *Random Structures and Algorithms*, to appear (also preprint math.PR/0312451 on <http://xxx.lanl.gov>), 2003.
- [3] R.W.R. Darling, J.R. Norris, Structure of large random hypergraphs, *Annals of Applied Probability*, to appear (also preprint math.PR/0109020 on <http://xxx.lanl.gov>), 2002.
- [4] C.A. Goldschmidt, Large random hypergraphs, Ph.D. thesis, University of Cambridge, 2003. Available from <http://www.statslab.cam.ac.uk/~cag27/>.
- [5] T.E. Harris, *The Theory of Branching Processes*, Springer-Verlag, 1963.
- [6] S. Janson, T. Luczak, A. Ruciński, *Random Graphs*, John Wiley and Sons, 2000.
- [7] B. Pittel, On tree census and the giant component in sparse random graphs, *Random Structures and Algorithms* **1** (1990), 311-342.